



US006645499B1

(12) **United States Patent**  
Lal et al.

(10) Patent No.: **US 6,645,499 B1**  
(45) Date of Patent: **\*Nov. 11, 2003**

(54) **HUMAN ENA/VASP-LIKE PROTEIN SPLICE VARIANT**

(75) Inventors: **Preeti Lal**, Santa Clara, CA (US); **Karl J. Guegler**, Menlo Park, CA (US); **Nell C. Corley**, Mountain View, CA (US)

(73) Assignee: **Incyte Corporation**, Palo Alto, CA (US)

(\*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **09/387,811**

(22) Filed: **Sep. 1, 1999**

#### Related U.S. Application Data

(62) Division of application No. 09/227,420, filed on Jan. 8, 1999, now Pat. No. 5,990,087, which is a division of application No. 09/026,587, filed on Feb. 20, 1998, now Pat. No. 5,912,128.

(51) Int. Cl.<sup>7</sup> ..... **A61K 39/00**

(52) U.S. Cl. .... **424/185.1; 530/350**

(58) Field of Search ..... **514/12; 530/350; 424/185.1**

(56) **References Cited**

#### PUBLICATIONS

Lerner, RA Tapping the immunological repertoire to produce antibodies of predetermined specificity. *Nature* 299:592-596, Oct. 1982.\*

Gertler, F.B. et al., "Mena, a Relative of VASP and Drosophila Enabled, Is Implicated in the Control of Microfilament Dynamics" *Cell* (1996) 87:227-239.

Ernekova, K.S. et al., "The WW Domain of Neural Protein FE65 Interacts with Proline-rich Motifs in Mena, the Mammalian Homolog of Drosophila Enabled" *J.Biol.Chem.* (1997) 272:32869-32877.

Linial, M., "Proline clustering in proteins from synaptic vesicles" *Neuroreport* (1994) 5:2009-2015.

Ohta, S. et al., "Differential Display Cloning of a Novel Rat cDNA (RNB6) That Shows High Expression in the Neonatal Brain Revealed a Member of Ena/VASP Family" *Biochem. Biophys. Res. Commun.* (1997) 237:307-312.

Gertler, F.B. et al., (GI 1644452), GenBank Sequence Database (Accession U72519), National Center for Biotechnology Information, National Library of Medicine, Bethesda, Maryland, 20894 (1996).

Walter, U., (GI 624963), GenBank Sequence Database (Accession Z46389), National Center for Biotechnology Information, National Library of Medicine, Bethesda, Maryland, 20894. (1995)

Haffner, C. et al., "Molecular cloning, structural analysis and functional expression of the proline-rich focal adhesion and microfilament-associated protein VASP" *EMBO J.* (1995) 14:19-27.

Adams, M.D., et al., GenBank Accession No. AA312190, cross reference to *Nature* 377, supp. 28:3-17, 1995.

Naeve, C.W., et al., Accuracy of Automated DNA sequencing: a multi-laboratory comparison of sequencing results. *Biotechniques*. 19(3):448-453. (1995).

Gertler, F.B., et al., GenBank Accession No. U72519, cross referenced to *Cell* 87(2):227-239 (1996).

Bork, *Genome Research*, vol. 10: 398-400, 2000.\*

Attwood, *Science* vol. 290: 471-473, 2000.\*

Skolnich et al. *TIBTECH* vol. 18: pp. 34-39, 2000.\*

\* cited by examiner

*Primary Examiner*—Patrick J. Nolan

(74) *Attorney, Agent, or Firm*—Incyte Corporation

(57) **ABSTRACT**

The invention provides a human ena/VASP-like protein splice variant (EVL1) and polynucleotides which identify and encode EVL1. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for treating or preventing disorders associated with expression of EVL1.

**2 Claims, 10 Drawing Sheets**

```

9      18      27      36      45      54
5' TTT AAG TAG GCT ATA AAA ATC AAG TTG CTG TCT TCA GAG GGT CTG TGG TCC TCT

63      72      81      90      99      108
GAT CAA CAT AGG CTG GTG GGA GTA CAG GAC TCG CCT CCG CAG GGT TCC CTG TGC

117      126      135      144      153      162
TGC CAC TTT TCA GCC ATG GCC ACA AGT GAA CAG AGT ATC TGC CAA GCC CGG GCT
M A T S E Q S I C Q A R A

171      180      189      198      207      216
TCC GTG ATG GTC TAC GAT GAC ACC AGT AAG AAA TGG GTA CCA ATC AAA CCT GGC
S V M V Y D D T S S K K W V P I K P G

225      234      243      252      261      270
CAG CAG GGA TTC AGC CGG ATC AAC ATC TAC CAC AAC ACT GCC AGC AAC ACC TTC
Q Q G F S R I N I Y H N T A S N T F

279      288      297      306      315      324
AGA GTC GTT GGA GTC AAG TTG CAG GAT CAG CAG GTT GTG ATC AAT TAT TCA ATC
R V V G V K L Q Q Q Q V V I N Y S I

333      342      351      360      369      378
GTG AAA GGG CTG AAG TAC AAT CAG GCC ACG CCA ACC TTC CAC CAG TGG CGA GAT
V K G L K Y N Q A T P T F H Q W R D

```

FIGURE 1A

387	GCC CGC CAG GTC TAC GGC TTA AAC TTT GCA AGT AAA GAA GAG GCA ACC ACG TTC	405	414	423	432
	A R Q	396	405	414	423
	V Y G L N F A S K E E A T T F				
441	TCC AAT GCA ATG CTG TTT GCC CTG AAC ATC ATG AAT TCC CAA GAA GGA GGC CCC	450	459	468	477
	S N A M L F A L N I M N S Q E G G P				
495	TCC AGC CAG CGT CAG GTG CAG AAT GGC CCC TCT CCT GAT GAG ATG GAC ATC CAG	504	513	522	531
	S Q R Q V Q N G P S P D E M D I Q				
549	AGA AGA CAA GTG ATG GAG CAG CAC CAG CAG CGT CAG GAA TCT CTA GAA AGA	558	567	576	585
	R Q V M E Q H Q Q Q R Q E S L E R				
603	AGA ACC TCG GCC ACA GGG CCC ATC CTC CCA CCA GGA CAT CCT TCA TCT GCA GCC	612	621	630	639
	R T S A T G P I L P P G H P S S A A				
657	AGC GCC CCC GTC TCA TGT AGT GGG CCT CCA CCG CCC CCC CCA CCT CTA GTC CCA	666	675	684	693
	S A P V S C S S G P P P P P P P L V P				
711	CCT CCA CCC ACT GGG GGT ACC CCA CCT CCC CCA CCC CCA CTG CCA GCC GGA GGA	720	729	738	747
	P P P T G A T P P P P P P L P A G G				

FIGURE 1B

765	GCC CAG GGG TCC AGC CAC GAG AGC TCC ATG TCA GGA CTG GCC GCT GCC ATA	774	783	792	801	810
A Q G S S H D E S S M S G L A A A I						
819	GCT GGG GCC AAG CTG AGA AGA GTC CAA CGG CCA GAA GAC GCA TCT GGA GGC TCC	828	837	846	855	864
A G A K L R R R V Q R P E D A S G S						
873	AGT CCC AGT GGG ACC TCA AAG TCC GAT GCC AAC CGG GCA AGC AGC GGG GGT GGC	882	891	900	909	918
S P S G T S S K S D A N R A S S G G						
927	GGA GGA GGC CTC ATG GAG GAA ATG AAC AAA CTG CTG GCC AAG AGG AGA AAA GCA	936	945	954	963	972
G G G L M E E M N K L L A K R R K A						
981	GCC TCC CAG TCA GAC AAG CCA GCC GAG AAG AAG GAA GAT GAA AGC CAA ATG GAA	990	999	1008	1017	1026
A S Q S D K P A E K K E D E S Q M E						
1035	GAT CCT AGT ACC TCC CCC TCT CCG GGG ACC CGA GCA GCC AGC CAG CCA CCT AAC	1044	1053	1062	1071	1080
D P S T S P S P G T R A A S Q P P N						
1089	TCC TCA GAG GCT GGC CGG AAG CCC TGG GAG CGG AGC AAC TCG GTG GAG AAG CCT	1098	1107	1116	1125	1134
S S E A G R K P W E R S N S V E K P						

FIGURE 1C

1143 1152 1161 1170 1179 1188  
GTG TCC TCG ATT CTG TCC AGA ACC CCG TCT GTG GCA AAG AGC CCC GAA GCT AAG  
V S S I L S R T P S V A K S P E A K

1197 1206 1215 1224 1233 1242  
AGC CCC CTT CAG TCG CAG CCT CAC TCT AGG ATG AAG CCT GGT GGG AGC GTG AAT  
S P L Q S Q P H S R M K P A G S V N

1251 1260 1269 1278 1287 1296  
GAC ATG GCC CTG GAT GCC TTC GAC TTG GAC CGG ATG AAG CAG GAG ATC CTA GAG  
D M A L D A F D L L D R M K Q E I L E

1305 1314 1323 1332 1341 1350  
GAG GTG GTG AGA GAG CTC CAC AAG GTG AAG GAG GAG ATC ATC GAC GCC ATC AGG  
E V V R E L H K V K E E I I D A I R

1359 1368 1377 1386 1395 1404  
CAG GAG CTG AGT GGG ATC AGC ACC ACG TAA GGG GCC GGC CTC GCT GCG CTG ATT  
Q E L S G I S T T

1413 1422 1431 1440 1449 1458  
CGT CGA GCC CAT CCG GCG ACA GAG GAC AGC CAG AAG CCC AGC CAG CCC CAG ACT

1467 1476 1485 1494 1503 1512  
CCA GTG CAC CAG AGC ACG CAC AGG AGC CTG GGC GCG CTG CTG TGA AAC GTC CTG

FIGURE 1D

1521	1530	1539	1548	1557	1566
ACC TGT GAT CAC ACA TGA CAG TGA GGA AAC CAA GTG CAA CTC CTG GGT TTT TTT					
1575	1584	1593	1602	1611	1620
TAG ATT CTG CCT GAC ACG GAA CAC CAG GTC TGC TCG TCT TTT TTT TGT TTT ATA					
1629	1638	1647	1656	1665	1674
TTT GCT TAT TTA AGG TAC ATT TCT TTG GGT TTC TAG AGA CGC CCC TAA GTC ACC					
1683	1692	1701	1710	1719	1728
TGC TTC ATT AGA CGG TTT CCA GGT TTT CTC CCA GGT GAC GCT GTT AGC GCC TCA					
1737	1746	1755	1764	1773	1782
GCT GGC GGT GAC AGC CGG CCC AGC GTG GCG CCA CCA CAC ACC GCA GAG CTG TCC					
1791	1800	1809	1818	1827	1836
AGG CAC AGC TCC GTC CCC AGC AGC GCT CAT GGT GTT GAA ACT GTC TGT CAT GCA CCA					
1845	1854	1863	1872	1881	
CGG TGT CTG TGT CCA CAC AGT AAT AAA CGG TTT ACT GTC CGC AAA AAA AA AA 3					

FIGURE 1E

1	M	A	T	S	E	Q	S	I	C	Q	A	R	A	S	V	M	V	Y	D	D	T	S	K	K	W	V	P	I	K	P	3089412
1	M	-	-	S	E	Q	S	I	C	Q	A	R	A	S	V	M	V	Y	D	D	T	S	K	K	W	V	P	I	K	P	GI 1644453
1	M	-	-	S	E	T	V	I	C	S	S	R	A	T	V	M	L	Y	D	D	G	N	K	R	W	L	P	A	G	T	GI 624964
31	G	Q	Q	G	F	S	R	I	N	I	Y	H	N	T	A	S	N	T	F	R	V	V	G	V	K	L	Q	-	D	Q	3089412
29	G	Q	Q	G	F	S	R	I	N	I	Y	H	N	T	A	S	S	T	F	R	V	V	G	V	K	L	Q	-	D	Q	GI 1644453
29	G	P	Q	A	F	S	R	V	Q	I	Y	H	N	P	T	A	N	S	F	R	V	V	G	R	K	M	Q	P	D	Q	GI 624964
60	Q	V	V	I	N	Y	S	I	V	K	G	L	K	Y	N	Q	A	T	P	T	F	H	Q	W	R	D	A	R	Q	V	3089412
58	Q	V	V	I	N	Y	S	I	V	K	G	L	K	Y	N	Q	A	T	P	T	F	H	Q	W	R	D	A	R	Q	V	GI 1644453
59	Q	V	V	I	N	C	A	I	V	R	G	V	K	Y	N	Q	A	T	P	N	F	H	Q	W	R	D	A	R	Q	V	GI 624964
90	Y	G	L	N	F	A	S	K	E	E	A	T	T	F	S	N	A	M	L	F	A	L	N	I	M	N	S	Q	E	G	3089412
88	Y	G	L	N	F	A	S	K	E	E	A	T	T	F	S	N	A	M	L	F	A	L	N	I	M	N	S	Q	E	G	GI 1644453
89	W	G	L	N	F	G	S	K	E	D	A	A	Q	F	A	A	G	M	A	S	A	L	E	A	L	E	G	G	P	GI 624964	
120	G	P	S	S	Q	-	-	-	R	Q	V	Q	N	G	P	S	P	D	E	M	D	I	Q	R	R	Q	V	M	E	Q	3089412
118	G	P	S	T	Q	-	-	-	R	Q	V	Q	N	G	P	S	P	E	E	M	D	I	Q	R	R	Q	V	M	E	Q	GI 1644453
119	P	P	P	A	L	P	T	W	S	V	P	N	G	P	S	P	E	E	V	E	Q	Q	K	R	Q	-	-	-	-	GI 624964	
147	H	Q	Q	R	Q	E	S	L	E	R	R	T	S	A	T	G	P	I	L	P	P	G	H	P	S	S	A	A	S	3089412	
145	-	-	-	Q	H	R	Q	E	S	L	E	R	R	I	S	A	T	G	P	I	L	P	P	G	H	P	S	S	A	S	GI 1644453
145	-	Q	P	G	P	S	E	H	I	E	R	R	V	S	N	A	G	G	-	-	P	P	A	P	P	A	G	G	P	P	GI 624964

FIGURE 2A

**FIGURE 2B**

**FIGURE 2B**



349	A	K	S	P	E	A	K	S	P	L	Q	S	Q	P	H	S	R	M	K	P	A	G	S	V	N	D	M	A	L	D	3089412
341	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	K	P	A	G	S	V	N	D	V	G	L	D	GI 1644453
321	-	K	S	S	S	V	T	T	S	E	T	Q	P	C	T	P	S	S	-	-	-	-	-	-	-	-	-	-	-	D	GI 624964
379	A	F	D	L	D	R	M	K	Q	E	I	L	E	E	V	V	R	E	L	H	K	V	K	E	E	I	I	D	A	I	3089412
354	A	L	D	L	D	R	M	K	Q	E	I	L	E	E	V	V	R	E	L	H	K	V	K	E	E	I	I	D	A	I	GI 1644453
341	Y	S	D	L	Q	R	V	K	Q	E	L	L	E	E	V	K	K	E	L	Q	K	V	K	E	E	I	I	E	A	F	GI 624964
409	R	Q	E	L	S	G	I	S	T	T																					3089412
384	R	Q	E	L	S	G	I	S	T	T																					GI 1644453
371	V	Q	E	L	R	K	R	G	S	P																					GI 624964

FIGURE 2C

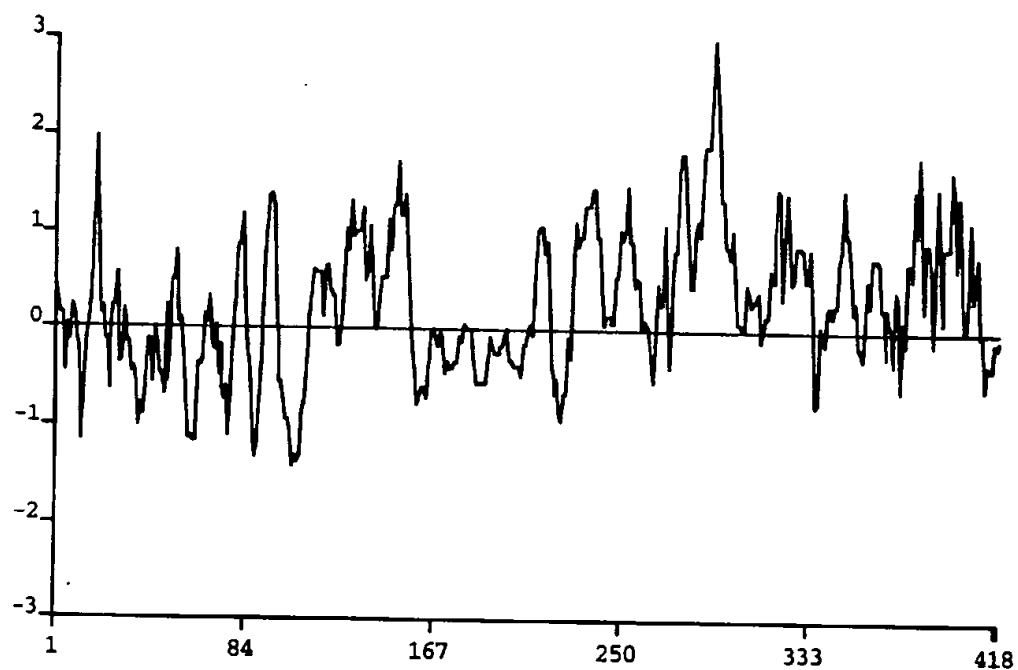


FIGURE 3A

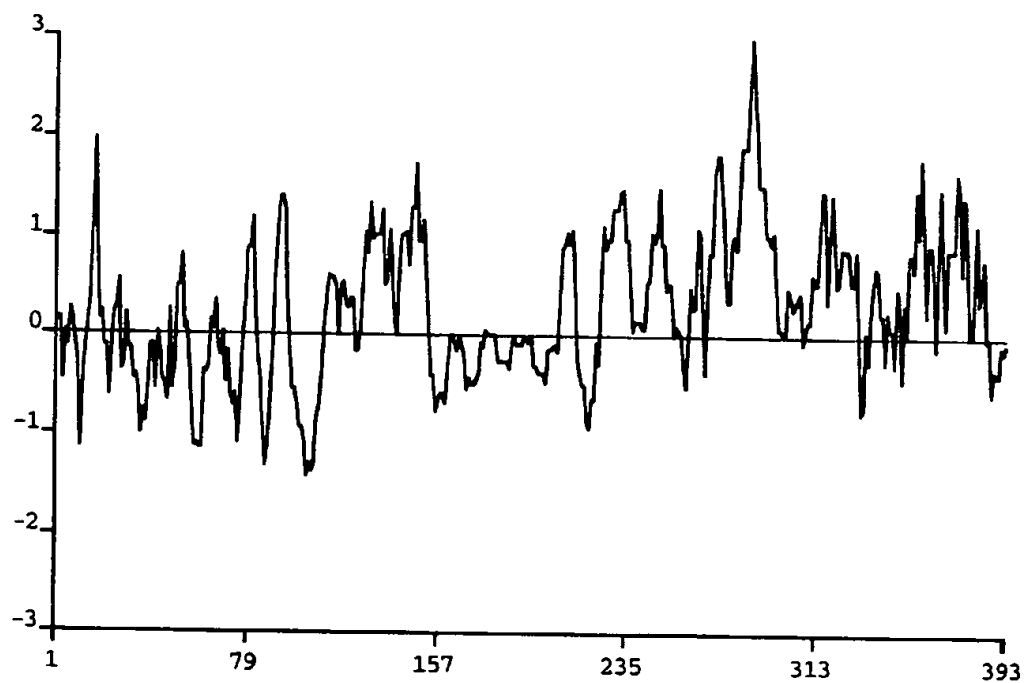


FIGURE 3B

## HUMAN ENA/VASP-LIKE PROTEIN SPLICE VARIANT

This application is a divisional application of U.S. application Ser. No. 09/227,420, filed Jan. 8, 1999, issued as U.S. Pat. No. 5,990,087 which is a divisional of U.S. application Ser. No. 09/026,587; filed Feb. 20, 1998, issued Jun. 15, 1999, as U.S. Pat. No. 5,912,128.

### FIELD OF THE INVENTION

This invention relates to nucleic acid and amino acid sequences of a human ena/VASP-like protein splice variant and to the use of these sequences in the diagnosis, treatment, and prevention of reproductive, immunological, vesicle trafficking, nervous system, developmental, and neoplastic disorders.

### BACKGROUND OF THE INVENTION

The control of cell morphology and motility requires the coupling of external stimuli to processes that alter the cytoskeletal architecture. The mechanical forces that drive morphological change and migration arise initially from the microfilament-based cytoskeleton. A large body of evidence links various signal transduction pathways to the formation of cellular outgrowths. The migration of neuronal growth cones is a well-studied mechanism for the actin-driven formation of membrane protrusions. In one example, the processes of axonal outgrowth are mediated by the *Drosophila* homolog of the c-Abl tyrosine kinase (Abl) and the product of the Disabled gene (Dab). Homozygous mutants of Abl and Dab make few or no proper axonal connections. The defects caused by loss of Abl and Dab in *Drosophila* are ameliorated by mutations in the Enabled (Ena) gene. Ena protein is tyrosine phosphorylated and has a proline-rich core which binds to the SH3 domains of Abl protein and Src protein *in vitro*. The murine homolog of Ena (Mena) and ena/VASP-like protein have recently been described and are members of a family of related molecules that include vasoactive-stimulated phosphoprotein (VASP). These proteins share three distinct regions of similarity: an amino-terminal 115 amino acids (EVH1 domain); a proline-rich core; and a carboxy-terminal 226 amino acids (EVH2 domain). Mena has phosphotyrosine and phosphoserine moieties and binds Abl and Src SH3 domains. (Gertler, F. B. et al. (1996) *Cell* 87:227-239.)

Human platelet activation is inhibited by agents such as prostaglandins and nitric oxide donors, which elevate intracellular cAMP or cGMP levels. Activation of platelets is associated with increased formation of intracellular F-actin. VASP is an abundant *in vivo* substrate for cyclic nucleotide-dependant protein kinases in platelets. VASP is a ligand for profilin, an actin-monomer binding protein that can stimulate the formation of F-actin. VASP is organized into three distinct domains. A central proline-rich domain contains a GPPPPP motif as a single copy and as a 3-fold tandem repeat, as well as three conserved phosphorylation sites for cyclic nucleotide-dependent protein kinases. A C-terminal domain contains a repetitive mixed-charge cluster which is predicted to form an alpha-helix. VASP expression in transiently transfected BHK21 cells was predominantly detected at stress fibers, at focal adhesions, and in F-actin-containing cell surface protrusions. In contrast, truncated VASP lacking the C-terminal domain was no longer concentrated at focal adhesions. These data indicate that the C-terminal domain is required for anchoring VASP at focal adhesion sites, while the central domain may mediate VASP interaction with

profilin. (Ermeikova, K. S. et al. (1997) *J. Biol. Chem.* 272:32869-32877.)

In comparison, Mena binds FE65, a neuronal protein which binds to the cytoplasmic portion of the  $\beta$ -amyloid precursor protein ( $\beta$ -APP).  $\beta$ -APP is a precursor to  $\beta$ -amyloid peptide, the major constituent of the extracellular plaques present in brain tissue from Alzheimer disease patients. Both VASP and Mena bind their respective adapter proteins (profilin or FE65) via distinct proline-rich regions and thus regulate adapter interaction(s) with other molecules. (Ermeikova, K. S. et al, *supra*.) Proline-rich domains and proline clusters have been identified in many proteins, particularly those that are associated with synaptic vesicles and other secretory organelles. These domains and clusters act as protein-protein interaction modules. (Linial, M. (1994) *Neuroreport* 5:2009-2015.) Two members of ena/VASP-like proteins have been recently isolated from mouse and rat and share 98.5% sequence identity. (Gertler, *supra*; and Ohta, S. et al. (1997) *Biochem. Biophys. Res. Comm.* 237:307-312.)

The discovery of a new human ena/VASP-like protein splice variant and the polynucleotides encoding it satisfies a need in the art by providing new compositions which are useful in the diagnosis, treatment, and prevention of reproductive, immunological, vesicle trafficking, nervous system, developmental, and neoplastic disorders.

### SUMMARY OF THE INVENTION

The invention is based on the discovery of a new human ena/VASP-like protein splice variant (EVL1), the polynucleotides encoding EVL1, and the use of these compositions for the diagnosis, treatment, or prevention of reproductive, immunological, vesicle trafficking, nervous system, developmental, and neoplastic disorders.

The invention features a substantially purified polypeptide, comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention further provides a substantially purified variant having at least 90% amino acid identity to the sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1. The invention also provides an isolated and purified polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1. The invention also includes an isolated and purified polynucleotide variant having at least 90% polynucleotide identity to the polynucleotide encoding the polypeptide consisting of the sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

Additionally, the invention provides a composition comprising a polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1. The invention further provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1, as well as an isolated and purified polynucleotide which is complementary to the polynucleotide encoding the polypeptide consisting of the sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides an isolated and purified polynucleotide comprising a sequence of SEQ ID NO:2 or a fragment of SEQ ID NO:2, and an isolated and purified polynucleotide variant having at least 90% polynucleotide identity to the polynucleotide comprising the sequence of SEQ ID NO:2 or a fragment of SEQ ID NO:2. The invention also provides an isolated and purified polynucleotide which

is complementary to the polynucleotide comprising the sequence of SEQ ID NO:2 or a fragment of SEQ ID NO:2.

The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1. In another aspect, the expression vector is contained within a host cell.

The invention also provides a method for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing at least a fragment of a polynucleotide encoding the polypeptide under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition consisting of a substantially purified polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1 in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide consisting of the sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1, as well as a purified agonist and a purified antagonist of the polypeptide.

The invention also provides a method for treating or preventing a reproductive disorder, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising substantially purified polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides a method for treating or preventing an immunological disorder, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising substantially purified polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides a method for treating or preventing a vesicle trafficking disorder, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising substantially purified polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides a method for treating or preventing a nervous system disorder, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising substantially purified polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides a method for treating or preventing a developmental disorder, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising substantially purified polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides a method for treating or preventing a neoplastic disorder, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides a method for detecting a polynucleotide encoding a polypeptide comprising the

amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1 in a biological sample containing nucleic acids, the method comprising the steps of: (a) hybridizing the complement of the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1 to at least one of the nucleic acids of the biological sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide encoding the polypeptide in the biological sample. In one aspect, the nucleic acids of the biological sample are amplified by the polymerase chain reaction prior to the hybridizing step.

#### BRIEF DESCRIPTION OF THE FIGS.

FIGS. 1A, 1B, 1C, 1D, and 1E show the amino acid sequence (SEQ ID NO:1) and nucleic acid sequence (SEQ ID NO:2) of EVL1. The alignment was produced using MACDMASIS PRO software (Hitachi Software Engineering Co. Ltd., San Bruno, Calif.).

FIGS. 2A, 2B, and 2C show the amino acid sequence alignments among EVL1 (3089412; SEQ ID NO:1), mouse ena/VASP-like protein (GI 1644453; SEQ ID NO:3), and human VASP (GI 624964; SEQ ID NO:4), produced using the multi sequence alignment program of DNASTAR software (DNASTAR Inc., Madison Wis.).

FIGS. 3A and 3B show the hydrophobicity plots for EVL1 (SEQ ID NO:1) and mouse ena/VASP-like protein (SEQ ID NO:3), respectively; the positive X axis reflects amino acid position and the negative Y axis reflects hydrophobicity (MACDMASIS PRO software).

#### DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## DEFINITIONS

"EVL1," as used herein, refers to the amino acid sequences of substantially purified EVL1 obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist," as used herein, refers to a molecule which, when bound to EVL1, increases or prolongs the duration of the effect of EVL1. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of EVL1.

An "allele" or an "allelic sequence," as these terms are used herein, is an alternative form of the gene encoding EVL1. Alleles may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding EVL1, as described herein, include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide the same EVL1 or a polypeptide with at least one functional characteristic of EVL1. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding EVL1, and improper or unexpected hybridization to alleles, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding EVL1. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent EVL1. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of EVL1 is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms "amino acid" or "amino acid sequence," as used herein, refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, "fragments," "immunogenic fragments," or "antigenic fragments" refer to fragments of EVL1 which are preferably about 5 to about 15 amino acids in length and which retain some biological activity or immunological activity of EVL1. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification," as used herein, relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction

(PCR) technologies well known in the art. (See, e.g., Dieffenbach, C. W. and G. S. Dveksler (1995) *PCR Primer, a Laboratory Manual*, Cold Spring Harbor Press, Plainview, N.Y., pp.1-5.)

The term "antagonist," as it is used herein, refers to a molecule which, when bound to EVL1, decreases the amount or the duration of the effect of the biological or immunological activity of EVL1. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of EVL1.

As used herein, the term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, F(ab)<sub>2</sub>, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind EVL1 polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant," as used herein, refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense," as used herein, refers to any composition containing a nucleic acid sequence which is complementary to a specific nucleic acid sequence. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

As used herein, the term "biologically active," refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic EVL1, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" or "complementarity," as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base pairing. For example, the sequence "A—G—T" binds to the complementary sequence "T—C—A." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding

between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" or a "composition consisting of a given amino acid sequence," as these terms are used herein, refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation, an aqueous solution, or a sterile composition. Compositions comprising polynucleotide sequences encoding EVL1 or fragments of EVL1 may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

The phrase "consensus sequence," as used herein, refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using XL-PCR (Perkin Elmer, Norwalk, Conn.) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW Fragment Assembly system (GCG, Madison, Wis.). Some sequences have been both extended and assembled to produce the consensus sequence.

As used herein, the term "correlates with expression of a polynucleotide" indicates that the detection of the presence of nucleic acids, the same or related to a nucleic acid sequence encoding EVL1, by northern analysis is indicative of the presence of nucleic acids encoding EVL1 in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding EVL1.

A "deletion," as the term is used herein, refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative," as used herein, refers to the chemical modification of EVL1, of a polynucleotide sequence encoding EVL1, or of a polynucleotide sequence complementary to a polynucleotide sequence encoding EVL1. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

The term "homology," as used herein, refers to a degree of complementarity. There may be partial homology or complete homology. The word "identity" may substitute for the word "homology." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially homologous sequence or hybridization probe will compete for and inhibit the binding of a completely homologous sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of

two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% homology or identity). In the absence of non-specific binding, the substantially homologous sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" or "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MEGALIGN program (Lasergene software package, DNASTAR, Inc., Madison, Wis.). The MEGALIGN program can create alignments between two or more sequences according to different methods, e.g., the clustal Method. (Higgins, D. G. and P. M. Sharp (1988) *Gene* 73:237-244.) The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no homology between the two amino acid sequences are not included in determining percentage similarity. Percent identity between nucleic acid sequences can also be calculated by the clustal method, or by other methods known in the art, such as the Jotun Hein Method. (See, e.g., Hein, J. (1990) *Methods in Enzymology* 183:626-645.) Identity between sequences can also be determined by other methods known in the art, e.g., by varying hybridization conditions.

"Human artificial chromosomes" (HACs), as described herein, are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance. (See, e.g., Harrington, J. J. et al. (1997) *Nat Genet.* 15:345-355.)

The term "humanized antibody," as used herein, refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization," as the term is used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

As used herein, the term "hybridization complex" as used herein, refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_{ot}$  or  $R_{ot}$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" or "addition," as used herein, refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, to the sequence found in the naturally occurring molecule.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be character-

ized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term "microarray," as used herein, refers to an array of distinct polynucleotides or oligonucleotides arrayed on a substrate, such as paper, nylon or any other type of membrane, filter, chip, glass slide, or any other suitable solid support.

The term "modulate," as it appears herein, refers to a change in the activity of EVL1. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of EVL1.

The phrases "nucleic acid" or "nucleic acid sequence," as used herein, refer to an oligonucleotide, nucleotide, polynucleotide, or any fragment thereof, to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material. In this context, "fragments" refers to those nucleic acid sequences which are greater than about 60 nucleotides in length, and most preferably are at least about 100 nucleotides, at least about 1000 nucleotides, or at least about 10,000 nucleotides in length.

The terms "operably associated" or "operably linked," as used herein, refer to functionally related nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in reading frame, certain genetic elements, e.g., repressor genes, are not contiguously linked to the encoded polypeptide but still bind to operator sequences that control expression of the polypeptide.

The term "oligonucleotide," as used herein, refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in a hybridization assay or microarray. As used herein, the term "oligonucleotide" is substantially equivalent to the terms "armpimers," "primers," "oligomers," and "probes," as these terms are commonly defined in the art.

"Peptide nucleic acid" (PNA), as used herein, refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA and RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell. (See, e.g., Nielsen, P. E. et al. (1993) *Anticancer Drug Des.* 8:53-63.)

The term "sample," as used herein, is used in its broadest sense. A biological sample suspected of containing nucleic acids encoding EVL1, or fragments thereof, or EVL1 itself may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a solid support; a tissue; a tissue print; etc.

As used herein, the terms "specific binding" or "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The interaction is dependent upon the presence of a particular structure of the protein recognized by the binding molecule (i.e., the antigenic determinant or epitope). For example, if

an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

As used herein, the term "stringent conditions" refers to conditions which permit hybridization between polynucleotide sequences and the claimed polynucleotide sequences. Suitably stringent conditions can be defined by, for example, the concentrations of salt or formamide in the prehybridization and hybridization solutions, or by the hybridization temperature, and are well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

For example, hybridization under high stringency conditions could occur in about 50% formamide at about 37° C. to 42° C. Hybridization could occur under reduced stringency conditions in about 35% to 25% formamide at about 30° C. to 35° C. In particular, hybridization could occur under high stringency conditions at 42° C. in 50% formamide, 5xSSPE, 0.3% SDS, and 200 µg/ml sheared and denatured salmon sperm DNA. Hybridization could occur under reduced stringency conditions as described above, but in 35% formamide at a reduced temperature of 35° C. The temperature range corresponding to a particular level of stringency can be further narrowed by calculating the purine to pyrimidine ratio of the nucleic acid of interest and adjusting the temperature accordingly. Variations on the above ranges and conditions are well known in the art.

The term "substantially purified," as used herein, refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution," as used herein, refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Transformation," as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, and refers to cells which transiently express the inserted DNA or RNA for limited periods of time.

A "variant" of EVL1, as used herein, refers to an amino acid sequence that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with isoleucine). More rarely, a variant may have "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNASTAR software.



## The Invention

The invention is based on the discovery of a new human ena/VASP-like protein splice variant (EVL1), the polynucleotides encoding EVL1, and the use of these compositions for the diagnosis, treatment, or prevention of reproductive, immunological, vesicle trafficking, nervous system, developmental, and neoplastic disorders.

Nucleic acids encoding the EVL1 of the present invention were first identified in Incyte Clone 3089412 from the aorta cDNA library (HEAONOT03) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:2, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 3089412 (HEAONOT03), 2836864 (TLYMNOT03), 1822064 (GBLATUT01), 1446806 (PLACNOT02), 1556238 (BLADTUT04), 1209813 (BRSTNOT02), and the shotgun sequence SAEA02787.

In one embodiment, the invention encompasses a polypeptide consisting of the amino acid sequence of SEQ ID NO:1, as shown in FIGS. 1A, 1B, 1C, 1D, and 1E. EVL1 is 418 amino acids in length and has two potential N-glycosylation sites at residues N64, and N319; one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at residue S160; eight potential casein kinase II phosphorylation sites at residues S96, S132, S215, S216, S253, S285, S298, and S371; six potential protein kinase C phosphorylation sites at residues T21, S22, T48, S123, T252, and S287; and a proline cluster from about residue P346 to about residue P368 and a predicted turn-coil-turn three-fold repeat structure from about residue T345 to about residue D374, indicative of a protein-protein interacting domain. As shown in FIGS. 2A, 2B, and 2C, EVL1 has chemical and structural homology with mouse ena/VASP-like protein (G1644453; SEQ ID NO:3), and human VASP (GI 624964; SEQ ID NO:4). In particular, EVL1 and mouse ena/VASP-like protein share 92% identity, two potential N-glycosylation sites, one potential cAMP- and cGMP-dependent protein kinase phosphorylation site, seven potential casein kinase II phosphorylation sites, and six potential protein kinase C phosphorylation sites. In addition, EVL1 and mouse ena/VASP-like protein have similar isoelectric points, 9.2 and 8.7, respectively. As illustrated by FIGS. 3A and 3B, EVL1 and mouse ena/VASP-like protein have rather similar hydrophobicity plots. The fragment of SEQ ID NO:2 from about nucleotide 1161 to about nucleotide 1197 is useful for designing oligonucleotides or to be used directly as a hybridization probe. Northern analysis shows the expression of this sequence in various libraries, at least 49% of which are immortalized or cancerous and at least 26% of which involve immune response. Of particular note is the expression of EVL1 in gastrointestinal, cardiovascular, neural, and developmental tissue; and in prostate, breast, ovary, and uterus tissue.

The invention also encompasses EVL1 variants. A preferred EVL1 variant is one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the EVL1 amino acid sequence, and which contains at least one functional or structural characteristic of EVL1.

The invention also encompasses polynucleotides which encode EVL1. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID NO:2, which encodes an EVL1.

The invention also encompasses a variant of a polynucleotide sequence encoding EVL1. In particular, such a variant polynucleotide sequence will have at least about 80%, more

preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding EVL1. A particular aspect of the invention encompasses a variant of SEQ ID NO:2 which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to SEQ ID NO:2. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of EVL1.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding EVL1, some bearing minimal homology to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring EVL1, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode EVL1 and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring EVL1 under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding EVL1 or its derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding EVL1 and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode EVL1 and EVL1 derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding EVL1 or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:2, or a fragment of SEQ ID NO:2, under various conditions of stringency. (See, e.g., Wahl, G. M. and S. L. Berger (1987) *Methods Enzymol.* 152:399-407; and Kimmel, A. R. (1987) *Methods Enzymol.* 152:507-511.)

Methods for DNA sequencing are well known and generally available in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical Corp., Cleveland, Ohio.), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, Ill.), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (GIBCO/BRL, Gaithersburg, Md.). Preferably, the process is automated with machines such as the microlab 2200 (Hamilton, Reno, Nev.), Peltier Thermal-Cycler (PTC200; MJ Research, Watertown, Mass.) and the ABI catalyst and 373 and 377 DNA sequencers (Perkin Elmer).

The nucleic acid sequences encoding EVL1 may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal primers to retrieve unknown sequence adjacent to a known locus. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.) In particular, genomic DNA is first amplified in the presence of a primer complementary to a linker sequence within the vector and a primer specific to the region predicted to encode the gene. The amplified sequences are then subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR may also be used to amplify or extend sequences using divergent primers based on a known region. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) The primers may be designed using commercially available software such as OLIGO 4.06 primer analysis software (National Biosciences Inc., Plymouth, Minn.) or another appropriate program to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68° C. to 72° C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Another method which may be used is capture PCR, which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to place an engineered double-stranded sequence into an unknown fragment of the DNA molecule before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J. D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060.) Additionally, one may use PCR, nested primers, and PROMOTER FINDER libraries to walk genomic DNA (Clontech, Palo Alto, Calif.). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable in that they will include more sequences which contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Perkin Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially pref-

erable for the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode EVL1 may be used in recombinant DNA molecules to direct expression of EVL1, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced, and these sequences may be used to clone and express EVL1.

As will be understood by those of skill in the art, it may be advantageous to produce EVL1-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce an RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter EVL1-encoding sequences for a variety of reasons including, but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding EVL1 may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of EVL1 activity, it may be useful to encode a chimeric EVL1 protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the EVL1 encoding sequence and the heterologous protein sequence, so that EVL1 may be cleaved and purified away from the heterologous moiety.

In another embodiment, sequences encoding EVL1 may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M. H. et al. (1980) *Nucl. Acids Res. Symp. Ser.* 215-223, and Hom, T. et al. (1980) *Nucl. Acids Res. Symp. Ser.* 225-232.) Alternatively, the protein itself may be produced using chemical methods to synthesize the amino acid sequence of EVL1, or a fragment thereof. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J. Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Perkin Elmer).

The newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R. M. and F. Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1983) *Proteins, Structures and Molecular Properties*, WH Freeman and Co., New York, N.Y.) Additionally, the amino acid sequence of EVL1, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a biologically active EVL1, the nucleotide sequences encoding EVL1 or derivatives thereof may be inserted into appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding EVL1 and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Press, Plainview, N.Y., ch. 4, 8, and 16-17; and Ausubel, F. M. et al. (1995, and periodic supplements) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, N.Y., ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding EVL1. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus (CaMV) or tobacco mosaic virus (TMV)) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

The "control elements" or "regulatory sequences" are those non-translated regions, e.g., enhancers, promoters, and 5' and 3' untranslated regions, of the vector and polynucleotide sequences encoding EVL1 which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters, e.g., hybrid lacZ promoter of the BLUESCRIPT phagemid (Stratagene, La Jolla, Calif.) or PSPORT1 plasmid (GIBCO/BRL), may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO, and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding EVL1, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for EVL1. For example, when large quantities of EVL1 are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, multi-functional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the sequence encoding EVL1 may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of  $\beta$ -galactosidase so that a hybrid protein is produced, and pIN vectors. (See, e.g., Van Hecke, G. and S. M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) pGEX vectors (Pharmacia Biotech, Uppsala, Sweden) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion

proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH, may be used. (See, e.g., Ausubel, supra; and Grant et al. (1987) *Methods Enzymol.* 153:516-544.) In cases where plant expression vectors are used, the expression of sequences encoding EVL1 may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV. (Takamatsu, N. (1987) *EMBO J.* 6:307-311.) Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews. (See, e.g., Hobbs, S. or Murry, L. E. in *McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill, New York, N.Y.; pp. 191-196.) An insect system may also be used to express EVL1. For example, in one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. The sequences encoding EVL1 may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of sequences encoding EVL1 will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, *S. frugiperda* cells or *Trichoplusia* larvae in which EVL1 may be expressed. (See, e.g., Engelhard, E. K. et al. (1994) *Proc. Nat. Acad. Sci.* 91:3224-3227.)

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding EVL1 may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing EVL1 in infected host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci.* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding EVL1. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding EVL1 and its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in

cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probi. Cell Differ.* 20:125-162.)

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding, and/or function. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and W138), are available from the American Type Culture Collection (ATCC, Bethesda, Md.) and may be chosen to ensure the correct modification and processing of the foreign protein.

For long term, high yield production of recombinant proteins, stable expression is preferred. For example, cell lines capable of stably expressing EVL1 can be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase genes and adenine phosphoribosyltransferase genes, which can be employed in tk or apr cells, respectively. (See, e.g., Wigler, M. et al. (1977) *Cell* 11:223-232; and Lowy, I. et al. (1980) *Cell* 22:817-823) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, dhfr, confers resistance to methotrexate; npt confers resistance to the aminoglycosides neomycin and G-418; and als or pat confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) *Proc. Natl. Acad. Sci.* 77:3567-3570; Colbere-Garapin, F. et al (1981) *J. Mol. Biol.* 150:1-14; and Murry, supra.) Additional selectable genes have been described, e.g., trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine. (See, e.g., Hartman, S. C. and R. C. Mulligan (1988) *Proc. Natl. Acad. Sci.* 85:8047-8051.) Recently, the use of visible markers has gained popularity with such markers as anthocyanins, B glucuronidase and its substrate GUS, luciferase and its substrate luciferin. Green fluorescent proteins (GFP) (Clontech, Palo Alto, Calif.) are also used (See, e.g., Chalfie, M. et al. (1994) *Science* 263:802-805.) These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C. A. et al. (1995) *Methods Mol. Biol.* 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding EVL1 is inserted within a marker gene sequence, transformed cells containing sequences encoding EVL1 can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding EVL1 under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells which contain the nucleic acid sequence encoding EVL1 and express EVL1 may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

The presence of polynucleotide sequences encoding EVL1 can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes or fragments or fragments of polynucleotides encoding EVL1. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences encoding EVL1 to detect transformants containing DNA or RNA encoding EVL1.

A variety of protocols for detecting and measuring the expression of EVL1, using either polyclonal or monoclonal antibodies specific for the protein, are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on EVL1 is preferred, but a competitive binding assay may be employed. These and other assays are well described in the art. (See, e.g., Hampton, R. et al. (1990) *Serological Methods, a Laboratory Manual*, APS Press, St Paul, Minn., Section IV; and Maddox, D. E. et al. (1983) *J. Exp. Med.* 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding EVL1 include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding EVL1, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Pharmacia & Upjohn (Kalamazoo, Minn.), Promega (Madison, Wis.), and U.S. Biochemical Corp. (Cleveland, Ohio.). Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding EVL1 may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The

protein produced by a transformed cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode EVL1 may be designed to contain signal sequences which direct secretion of EVL1 through a prokaryotic or eukaryotic cell membrane. Other constructions may be used to join sequences encoding EVL1 to nucleotide sequences encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAG extension/affinity purification system (Immunex Corp., Seattle, Wash.). The inclusion of cleavable linker sequences, such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, Calif.), between the purification domain and the EVL1 encoding sequence may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing EVL1 and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on immobilized metal ion affinity chromatography. (IMAC) (See, e.g., Porath, J. et al. (1992) *Prot. Exp. Purif.* 3: 263-281.) The enterokinase cleavage site provides a means for purifying EVL1 from the fusion protein. (See, e.g., Kroll, D. J. et al. (1993) *DNA Cell Biol.* 12:441-453.) Fragments of EVL1 may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, T.E. (1984) *Protein: Structures and Molecular Properties*, pp. 55-60, W. H. Freeman and Co., New York, N.Y.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the Applied Biosystems 431 A peptide synthesizer (Perkin Elmer). Various fragments of EVL1 may be synthesized separately and then combined to produce the full length molecule.

#### Therapeutics

Chemical and structural homology exists among EVL1, mouse ena/VASP-like protein (GI 1644453), and human VASP (GI 624964). In addition, EVL1 is expressed in cancer and the immune response; in gastrointestinal, cardiovascular, neural, and developmental tissue; and in prostate, breast, ovary, and uterus tissue. Therefore, EVL1 appears to play a role in reproductive, immunological, vesicle trafficking, nervous system, developmental, and neoplastic disorders.

Therefore, in one embodiment, EVL1 or a fragment or derivative thereof may be administered to a subject to treat or prevent a reproductive disorder. Such reproductive disorders can include, but are not limited to, disorders of prolactin production; infertility, including tubal disease, ovulatory defects, and endometriosis; disruptions of the estrous cycle, disruptions of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, autoimmune disorders, ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, and prostatitis, carcinoma of the male breast and gynecomastia.

In another embodiment, a vector capable of expressing EVL1 or a fragment or derivative thereof may be adminis-

tered to a subject to treat or prevent a reproductive disorder including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified EVL1 in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a reproductive disorder including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of EVL1 may be administered to a subject to treat or prevent a reproductive disorder including, but not limited to, those listed above.

In another embodiment, EVL1 or a fragment or derivative thereof may be administered to a subject to treat or prevent an immunological disorder. Such immunological disorders can include, but are not limited to, AIDS, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hyper eosinophilia, irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjogren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, ulcerative colitis, Werner syndrome, and complications of cancer, hemodialysis, and extracorporeal circulation; viral, bacterial, fungal, parasitic, protozoal, and helminthic infections; and trauma.

In another embodiment, a vector capable of expressing EVL1 or a fragment or derivative thereof may be administered to a subject to treat or prevent an immunological disorder including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified EVL1 in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent an immunological disorder including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of EVL1 may be administered to a subject to treat or prevent an immunological disorder including, but not limited to, those listed above.

In another embodiment, EVL1 or a fragment or derivative thereof may be administered to a subject to treat or prevent a vesicle trafficking disorder. Such vesicle trafficking disorders can include, but are not limited to, cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, diabetes mellitus, diabetes insipidus, hyper- and hypoglycemia, Grave's disease, goiter, Cushing's disease, and Addison's disease; gastrointestinal disorders including ulcerative colitis, gastric and duodenal ulcers; other conditions associated with abnormal vesicle trafficking including AIDS; allergies including hay fever, asthma, and urticaria (hives); autoimmune hemolytic anemia; proliferative glomerulonephritis; inflammatory bowel disease; multiple sclerosis; myasthenia gravis; rheumatoid and osteoarthritis; scleroderma; Chediak-Higashi and Sjogren's syndromes; systemic lupus erythematosus; toxic shock syndrome; traumatic tissue damage; and viral, bacterial, fungal, helminth, and protozoal infections.

In another embodiment, a vector capable of expressing EVL1 or a fragment or derivative thereof may be administered to a subject to treat or prevent a vesicle trafficking disorder including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified EVL1 in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a vesicle trafficking disorder including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of EVL1 may be administered to a subject to treat or prevent a vesicle trafficking disorder including, but not limited to, those listed above.

In another embodiment, EVL1 or a fragment or derivative thereof may be administered to a subject to treat or prevent a nervous system disorder. Such nervous system disorders can include, but are not limited to, akathisia, Alzheimer's disease, amnesia, amyotrophic lateral sclerosis, bipolar disorder, catatonia, cerebral neoplasms, dementia, depression, diabetic neuropathy, Down's syndrome, tardive dyskinesia, dystonias, epilepsy, Huntington's disease, multiple sclerosis, neurofibromatosis, Parkinson's disease, paranoid psychoses, postherpetic neuralgia, schizophrenia, and Tourette's disorder; angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, and pheochromocytoma.

In another embodiment, a vector capable of expressing EVL1 or a fragment or derivative thereof may be administered to a subject to treat or prevent a nervous system disorder including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified EVL1 in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a nervous system disorder including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of EVL1 may be administered to a subject to treat or prevent a nervous system disorder including, but not limited to, those listed above.

In another embodiment, EVL1 or a fragment or derivative thereof may be administered to a subject to treat or prevent a developmental disorder. The term "developmental disorder" refers to any disorder associated with development or function of a tissue, organ, or system of a subject (such as the brain, adrenal gland, kidney, skeletal or reproductive system). Such developmental disorders can include, but are not limited to, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome, Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spinal bifida, and congenital glaucoma, cataract, or sensorineural hearing loss.

In another embodiment, a vector capable of expressing EVL1 or a fragment or derivative thereof may be administered to a subject to treat or prevent a developmental disorder including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified EVL1 in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a developmental disorder including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of EVL1 may be administered to a subject to treat or prevent a developmental disorder including, but not limited to, those listed above.

In a further embodiment, an antagonist of EVL1 may be administered to a subject to treat or prevent a neoplastic disorder. Such a neoplastic disorder may include, but is not limited to, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In one aspect, an antibody which specifically binds EVL1 may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express EVL1.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding EVL1 may be administered to a subject to treat or prevent a neoplastic disorder including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of EVL1 may be produced using methods which are generally known in the art. In particular, purified EVL1 may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind EVL1. Antibodies to EVL1 may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with EVL1 or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to EVL1 have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of EVL1 amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to EVL1 may be prepared using any technique which provides for the production of antibody

molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R. J. et al. (1983) *Proc. Natl. Acad. Sci.* 80:2026-2030; and Cole, S. P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S. L. et al. (1984) *Proc. Natl. Acad. Sci.* 81:6851-6855; Neuberger, M. S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce EVL1-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotype composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton D. R. (1991) *Proc. Natl. Acad. Sci.* 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci.* 86: 3833-3837; and Winter, G. et al. (1991) *Nature* 349:293-299.)

Antibody fragments which contain specific binding sites for EVL1 may also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W. D. et al. (1989) *Science* 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between EVL1 and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering EVL1 epitopes is preferred, but a competitive binding assay may also be employed. (Maddox, supra.)

In another embodiment of the invention, the polynucleotides encoding EVL1, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding EVL1 may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding EVL1. Thus, complementary molecules or fragments may be used to modulate EVL1 activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding EVL1.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide

sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors which will express nucleic acid sequences complementary to the polynucleotides of the gene encoding EVL1. (See, e.g., Sambrook, supra; and Ausubel, supra.)

Genes encoding EVL1 can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding EVL1. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding EVL1. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J. E. et al. (1994) in Huber, B. E. and B. I. Carr, *Molecular and Immunologic Approaches*, Futura Publishing Co., Mt. Kisco, N.Y., pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding EVL1.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated in vitro and in vivo transcription of DNA sequences encoding EVL1. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.



RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C. K. et al. (1997) *Nature Biotechnology* 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of EVL1, antibodies to EVL1, and mimetics, agonists, antagonists, or inhibitors of EVL1. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intrarterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of *Remington's Pharmaceutical Sciences* (Maack Publishing Co., Easton, Pa.).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch

from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acids. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of EVL1, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredi-



ents are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example EVL1 or fragments thereof, antibodies of EVL1, and agonists, antagonists or inhibitors of EVL1, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED50 (the dose therapeutically effective in 50% of the population) or LD50 (the dose lethal to 50% of the population) statistics. The dose ratio of therapeutic to toxic effects is the therapeutic index, which can be expressed as the ED50/LD50 ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED50 with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu$ g to 100,000  $\mu$ g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

#### Diagnostics

In another embodiment, antibodies which specifically bind EVL1 may be used for the diagnosis of disorders characterized by expression of EVL1, or in assays to monitor patients being treated with EVL1 or agonists, antagonists, or inhibitors of EVL1. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for EVL1 include methods which utilize the antibody and a label to detect EVL1 in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring EVL1, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of EVL1 expression. Normal or standard values for EVL1 expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with is antibody to EVL1 under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of EVL1 expressed in subject, samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding EVL1 may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of EVL1 may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of EVL1, and to monitor regulation of EVL1 levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding EVL1 or closely related molecules may be used to identify nucleic acid sequences which encode EVL1. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies only naturally occurring sequences encoding EVL1, alleles, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably contain at least 50% of the nucleotides from any of the EVL1 encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:2 or from genomic sequences including promoters, enhancers, and introns of the EVL1 gene.

Means for producing specific hybridization probes for DNAs encoding EVL1 include the cloning of polynucleotide sequences encoding EVL1 or EVL1 derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}$ P or  $^{35}$ S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding EVL1 may be used for the diagnosis of a disorder associated with expression of EVL1. Examples of such a disorder include, but are not limited to, a reproductive disorder, such as, disorders of prolactin production; infertility, including tubal disease, ovulatory defects, and endometriosis; disruptions of the estrous cycle, disruptions of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, autoimmune disorders, ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia,

and prostatitis, carcinoma of the male breast and gynecomasia; an immunological disorder, such as, AIDS, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, ulcerative colitis, Werner syndrome, and complications of cancer, hemodialysis, and extracorporeal circulation; viral, bacterial, fungal, parasitic, protozoal, and helminthic infections; and trauma; a vesicle trafficking disorder, such as, cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, diabetes mellitus, diabetes insipidus, hyper- and hypoglycemia, Grave's disease, goiter, Cushing's disease, and Addison's disease; gastrointestinal disorders including ulcerative colitis, gastric and duodenal ulcers; other conditions associated with abnormal vesicle trafficking including AIDS; allergies including hay fever, asthma, and urticaria (hives); autoimmune hemolytic anemia; proliferative glomerulonephritis; inflammatory bowel disease; multiple sclerosis; myasthenia gravis; rheumatoid and osteoarthritis; scleroderma; Chediak-Higashi and Sjögren's syndromes; systemic lupus erythematosus; toxic shock syndrome; traumatic tissue damage; and viral, bacterial, fungal, helminth, and protozoal infections; a nervous system disorder, such as, akathisia, Alzheimer's disease, amnesia, amyotrophic lateral sclerosis, bipolar disorder, catatonia, cerebral neoplasms, dementia, depression, diabetic neuropathy, Down's syndrome, tardive dyskinesia, dystonias, epilepsy, Huntington's disease, multiple sclerosis, neurofibromatosis, Parkinson's disease, paranoid psychoses, postherpetic neuralgia, schizophrenia, and Tourette's disorder; angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, and pheochromocytoma; a developmental disorder, such as, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome, Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucocutaneous dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spinal bifida, and congenital glaucoma, cataract, or sensorineural hearing loss, and a neoplastic disorder, such as, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding EVL1 may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and ELISA assays; and in microarrays utilizing fluids or tissues from patients to detect altered EVL1 expres-

sion. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding EVL1 may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding EVL1 may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding EVL1 in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of EVL1, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding EVL1, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding EVL1 may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding EVL1, or a fragment of a polynucleotide complementary to the polynucleotide encoding EVL1, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of EVL1 include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P. C. et al. (1993) *J. Immunol. Methods* 159:235-244; and Duplaa, C. et al. (1993) *Anal. Biochem* 212:229-236.) The

speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T. M. et al. (1995) U.S. Pat. No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R. A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M. J. et al. (1997) U.S. Pat. No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding EVL1 may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial plasmid constructions, or single chromosome cDNA libraries. (See, e.g., Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B. J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, R. A. (ed.) *Molecular Biology and Biotechnology*, VCH Publishers New York, N.Y., pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding EVL1 on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., AT to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R. A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect

differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, EVL1, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between EVL1 and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application W084/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with EVL1, or fragments thereof, and washed. Bound EVL1 is then detected by methods well known in the art. Purified EVL1 can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding EVL1 specifically compete with a test compound for binding EVL1. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with EVL1.

In additional embodiments, the nucleotide sequences which encode EVL1 may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

## EXAMPLES

### I. HEAONOT03 cDNA Library Construction

The HEAONOT03 cDNA library was constructed from normal aorta tissue obtained from a 27-year-old Caucasian female who died from intracranial bleeding.

The frozen tissue was homogenized and lysed TRIZOL reagent (1 g tissue/10 ml TRIZOL, Catalog #10296-028; GIBCO-BRL), a monophasic solution of phenol and guanidine thiocyanate, using a Polytron PT-3000 homogenizer (Brinkmann Instruments, Westbury, N.Y.). After a brief incubation on ice, chloroform was added (1:5 v/v) and the lysate was centrifuged. The upper aqueous layer was removed to a fresh tube and the RNA precipitated with isopropanol, resuspended in DEPC-treated water, and treated with DNase for 25 min at 37° C. The RNA was extracted and precipitated as described before. The mRNA was then isolated using the OLIGOTEX (QIAGEN, Inc., Chatsworth, Calif.) and used to construct the cDNA library.

The mRNA was handled according to the recommended protocols in the SUPERScript plasmid system (Catalog #18248-013, GIBCO-BRL). cDNA synthesis was initiated with a Not I-oligo d(T) primer. Double-stranded cDNA was blunted, ligated to EcoR I adaptors, digested with Not I, fractionated on a SEPHAROSE CL4B column (Catalog

#275105-01; Pharmacia), and those cDNAs exceeding 400 bp were ligated into the Not I- and EcoR I sites of the pINCY 1 vector (Incyte). The plasmid pINCY 1 was subsequently transformed into DH5 $\alpha$  competent cells (Catalog #18258-012; GIBCO-BRL).

## II. Isolation and Sequencing of cDNA Clones

Plasmid DNA was released from the cells and purified using the REAL PREP 96 plasmid kit (Catalog #26173; QIAGEN, Inc.). The recommended protocol was employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile Terrific Broth (Catalog #22711, GIBCO-BRL) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) after inoculation, the cultures were incubated for 19 hours and at the end of incubation, the cells were lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellet was resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples were transferred to a 96-well block for storage at 4° C.

The cDNAs were sequenced by the method of Sanger et al. (1975, *J. Mol. Biol.* 94:441f), using a MICROLAB 2200 (Hamilton, Reno, Nev.) in combination with Peltier thermal cyclers (PTC200 from MJ Research, Watertown, Mass.) and Applied Biosystems 377 DNA sequencing systems, and the reading frame was determined.

## III. Homology Searching of cDNA Clones and Their Deduced Proteins

The nucleotide sequences and/or amino acid sequences of the Sequence Listing were used to query sequences in the GenBank, SwissProt, BLOCKS, and Pima II databases. These databases, which contain previously identified and annotated sequences, were searched for regions of homology using BLAST (Basic Local Alignment Search Tool). (See, e.g., Altschul, S. F. (1993) *J. Mol. Evol.* 36:290-300; and Altschul et al. (1990) *J. Mol. Biol.* 215:403-410.)

BLAST produced alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST was especially useful in determining exact matches or in identifying homologs which may be of prokaryotic (bacterial) or eukaryotic (animal, fungal, or plant) origin. Other algorithms could have been used when dealing with primary sequence patterns and secondary structure gap penalties. (See, e.g., Smith, T. et al. (1992) *Protein Engineering* 5:35-51.) The sequences disclosed in this application have lengths of at least 49 nucleotides and have no more than 12% uncalled bases (where N is recorded rather than A, C, G, or T).

The BLAST approach searched for matches between a query sequence and a database sequence. BLAST evaluated the statistical significance of any matches found, and reported only those matches that satisfy the user-selected threshold of significance. In this application, threshold was set at 10<sup>-23</sup> for nucleotides and 10<sup>-8</sup> for peptides.

Incyte nucleotide sequences were searched against the GenBank databases for primate PF-0471 US (pri), rodent (rod), and other mammalian sequences (mam), produced amino acid sequences from the same clones were then searched against GenBank functional protein databases, mammalian (mamp), vertebrate (vrtp), and eukaryote (eukp), for homology.

Additionally, sequences identified from cDNA libraries may be analyzed to identify those gene sequences encoding

conserved protein motifs using an appropriate analysis program, e.g., the Block 2 Bioanalysis Program (Incyte, Palo Alto, Calif.). This motif analysis program, based on sequence information contained in the Swiss-Prot-Database and PROSITE, is a method of determining the function of uncharacterized proteins translated from genomic or cDNA sequences. (See, e.g., Bairoch, A. et al. (1997) *Nucleic Acids Res.* 25:217-221; and Attwood, T. K. et al. (1997) *J. Chem. Inf. Comput. Sci.* 37:417424.) PROSITE may be used to identify common functional or structural domains in divergent proteins. The method is based on weight matrices. Motifs identified by this method are then calibrated against the SWISS-PROT database in order to obtain a measure of the chance distribution of the matches.

In another alternative, Hidden Markov models (HMMs) may be used to find protein domains, each defined by a dataset of proteins known to have a common biological function. (See, e.g., Pearson, W. R. and D. J. Lipman (1988) *Proc. Natl. Acad. Sci.* 85:2444-2448; and Smith, T. F. and M. S. Waterman (1981) *J. Mol. Biol.* 147:195-197.) HMMs were initially developed to examine speech recognition patterns, but are now being used in a biological context to analyze protein and nucleic acid sequences as well as to model protein structure. (See, e.g., Krogh, A. et al. (1994) *J. Mol. Biol.* 235:1501-1531; and Collin, M. et al. (1993) *Protein Sci.* 2:305-314.) HMMs have a formal probabilistic basis and use position-specific scores for amino acids or nucleotides. The algorithm continues to incorporate information from newly identified sequences to increase its motif analysis capabilities.

## IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; and Ausubel, F. M. et al. supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST are used to search for identical or related molecules in nucleotide databases such as GENBANK or LIFESEQ database (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous.

The basis of the search is the product score, which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Homologous molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analysis are reported as a list of libraries in which the transcript encoding EVL1 occurs. Abundance and percent abundance are also reported. Abundance directly reflects the number of times a particular transcript is represented in a cDNA library, and percent abundance is abundance divided by the total number of sequences examined in the cDNA library.

## V. Extension of EVL1 Encoding Polynucleotides

The nucleic acid sequence of Incyte Clone 3089412 was used to design oligonucleotide primers for extending a partial nucleotide sequence to full length. One primer was synthesized to initiate extension of an antisense polynucleotide, and the other was synthesized to initiate extension of a sense polynucleotide. Primers were used to facilitate the extension of the known sequence "outward" generating amplicons containing new unknown nucleotide sequence for the region of interest. The initial primers were designed from the cDNA using OLIGO 4.06 software (National Biosciences, Plymouth, Minn.), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68° C. to about 72° C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries (GIBCO/BRL) were used to extend the sequence. If more than one extension is necessary or desired, additional sets of primers are designed to further extend the known region.

High fidelity amplification was obtained by following the instructions for the XL-PCR kit (Perkin Elmer) and thoroughly mixing the enzyme and reaction mix. PCR was performed using the Peltier thermal cycler (PTC200; M. J. Research, Watertown, Mass.), beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, with the following parameters:

Step 1	94° C. for 1 min (initial denaturation)
Step 2	65° C. for 1 min
Step 3	68° C. for 6 min
Step 4	94° C. for 15 sec
Step 5	65° C. for 1 min
Step 6	68° C. for 7 min
Step 7	Repeat steps 4 through 6 for an additional 15 cycles
Step 8	94° C. for 15 sec
Step 9	65° C. for 1 min
Step 10	68° C. for 7:15 min
Step 11	Repeat steps 8 through 10 for an additional 12 cycles
Step 12	72° C. for 8 min
Step 13	4° C. (and holding)

A 5  $\mu$ l to 10  $\mu$ l aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration (about 0.6% to 0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products were excised from the gel, purified using QIAQUICK (QIAGEN Inc., Chatsworth, Calif.), and trimmed of overhangs using Klenow enzyme to facilitate religation and cloning.

After ethanol precipitation, the products were redissolved in 13  $\mu$ l of ligation buffer, 1  $\mu$ l T4-DNA ligase (15 units) and 1  $\mu$ l T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2 to 3 hours, or overnight at 16° C. Competent *E. coli* cells (in 40  $\mu$ l of appropriate media) were transformed with 3  $\mu$ l of ligation mixture and cultured in 80  $\mu$ l of SOC medium. (See, e.g., Sambrook, supra, Appendix A, p. 2.) After incubation for one hour at 37° C., the *E. coli* mixture was plated on Luria Bertani (LB) agar (See, e.g., Sambrook, supra, Appendix A, p. 1) containing 2xCarb. The following day, several colonies were randomly picked from each plate and cultured in 150  $\mu$ l of liquid LB/2xCarb medium placed in an individual well of an appropriate commercially-available sterile 96-well

microtiter plate. The following day, 5  $\mu$ l of each overnight culture was transferred into a non-sterile 96-well plate and, after dilution 1:10 with water, 5  $\mu$ l from each sample was transferred into a PCR array.

For PCR amplification, 18  $\mu$ l of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer, and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using the following conditions:

Step 1	94° C. for 60 sec
Step 2	94° C. for 20 sec
Step 3	55° C. for 30 sec
Step 4	72° C. for 90 sec
Step 5	Repeat steps 2 through 4 for an additional 29 cycles
Step 6	72° C. for 180 sec
Step 7	4° C. (and holding)

Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid, and sequenced.

In like manner, the nucleotide sequence of SEQ ID NO:2 is used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for 5' extension, and an appropriate genomic library.

## VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:2 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham, Chicago, Ill.), and T4 polynucleotide kinase (DuPont NEN, Boston, Mass.). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine resin column (Pharmacia & Upjohn, Kalamazoo, Mich.). An aliquot containing 10<sup>7</sup> counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN, Boston, Mass.).

The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham, N.H.). Hybridization is carried out for 16 hours at 40° C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMATAR film (Kodak, Rochester, N.Y.) is exposed to the blots, hybridization patterns are compared visually.

## VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV,

chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE. Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) *Science* 270:467-470; and Shalon, D. et al. (1996) *Genome Res.* 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

#### VIII. Complementary Polynucleotides

Sequences complementary to the EVL1-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring EVL1. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software and the coding sequence of EVL1. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the EVL1-encoding transcript.

#### IX. Expression of EVL1

Expression of EVL1 is accomplished by subcloning the cDNA into an appropriate vector and transforming the vector into host cells. This vector contains an appropriate promoter, e.g.,  $\beta$ -galactosidase upstream of the cloning site, operably associated with the cDNA of interest. (See, e.g., Sambrook, supra pp. 404433; and Rosenberg, M. et al. (1983) *Methods Enzymol.* 101:123-138.)

Induction of an isolated, transformed bacterial strain with isopropyl beta-D-thiogalactopyranoside (IPTG) using standard methods produces a fusion protein which consists of the first 8 residues of  $\beta$ -galactosidase, about 5 to 15 residues of linker, and the full length protein. The signal residues direct the secretion of EVL1 into bacterial growth media which can be used directly in the following assay for activity.

#### X. Demonstration of EVL1 Activity

The assay for human ena/VASP-like protein splice variant (EVL1) is based upon the binding affinity of mouse ena/VASP for the neural protein FE65. (Ermeikova, K. S. et al. supra.) Monkey cell line COS-7 (ATCC CRL 1651) cells are grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin mixture in 5% CO<sub>2</sub> atmosphere at 37° C.

3×10<sup>6</sup> cells are transfected by electroporation with 20  $\mu$ g of CMV-hemagglutinin-FE65 plasmid and 20  $\mu$ g of CMV-EVL1 plasmid as known in the art. 62 h after the transfection, the cells are harvested in ice-cold PBS and centrifuged at 2000 rpm at 4° C., and the pellet is dissolved in lysis buffer (10 mM Tris HCl, pH 7.5; 150 mM NaCl; 0.1 mM sodium vanadate; 50 mM NaF; 0.5% Nonidet P-40; 1 mM phenylmethylsulfonyl fluoride; 10  $\mu$ g/ml each of aprotinin, leupeptin, and pepstatin). The extracts are clarified by centrifugation at 16,000 g at 4° C., and 4 mg of supernatant are incubated for 1 h at 4° C. with an anti-hemagglutinin monoclonal antibody or with an unrelated monoclonal antibody. Thereafter, 30  $\mu$ l of protein A-SEPHAROSE resin (Pharmacia) are added to each sample of the extract-antibody mixture, and the immunocomplexes are eluted with 50 mM Tris HCl, pH 6.8, 2% SDS, 10% glycerol, 100 mM dithiothreitol, 0.01% bromophenol blue. The proteins are resolved by 7.5% SDS-PAGE and transferred to Immobilon-P membranes (Millipore). The filter is blocked in 5% nonfat dry milk in tris-buffered saline, 0.5 % Tween, buffer (TBS-T) and incubated with anti-EVL1 antibodies at 1:1000 dilution for 1 h at room temperature. After washing in TBS-T, the filter is exposed to horseradish peroxidase-conjugated protein A (Amersham Corp.) at a dilution of 1:5000 for 30 min at room temperature. The signals are detected by chemiluminescence using the ECL system (Amersham Corp.). The signal response is proportional to the activity of EVL1 in the preparation.

#### XI. Production of EVL1 Specific Antibodies

EVL1 substantially purified using PAGE electrophoresis (see, e.g., Harrington, M. G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols. The EVL1 amino acid sequence is analyzed using DNASTAR software (DNASTAR Inc) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel et al., ch. 11.)

Typically, the oligopeptides are 15 residues in length, and are synthesized using an Applied Biosystems 431A Peptide Synthesizer using fmoc-chemistry and coupled to KLH (Sigma, St. Louis, Mo.) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel et al. supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity, for example, by binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### XII. Purification of Naturally Occurring EVL1 Using Specific Antibodies

Naturally occurring or recombinant EVL1 is substantially purified by immunoaffinity chromatography using antibodies specific for EVL1. An immunoaffinity column is constructed by covalently coupling anti-EVL1 antibody to an activated chromatographic resin, such as CNBr-activated Sepharose (Pharmacia & Upjohn). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing EVL1 are passed over the immunoaffinity column, and the column is washed under conditions

that allow the preferential absorbance of EVL1 (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/EVL1 binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and EVL1 is collected.

### XIII. Identification of Molecules Which Interact with EVL1

EVL1, or biologically active fragments thereof, are labeled with  $^{125}\text{I}$  Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled EVL1, washed, and any wells with labeled EVL1 complex are assayed. Data obtained

using different concentrations of EVL1 are used to calculate values for the number, affinity, and association of EVL1 with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

#### SEQUENCE LISTING

##### (1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 4

##### (2) INFORMATION FOR SEQ ID NO:1:

###### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 418 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

###### (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: HEAONOT03
- (B) CLONE: 3089412

###### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```

Met Ala Thr Ser Glu Gln Ser Ile Cys Gln Ala Arg Ala Ser Val Met
 1           5           10           15
Val Tyr Asp Asp Thr Ser Lys Lys Trp Val Pro Ile Lys Pro Gly Gln
20           25           30
Gln Gly Phe Ser Arg Ile Asn Ile Tyr His Asn Thr Ala Ser Asn Thr
35           40           45
Phe Arg Val Val Gly Val Lys Leu Gln Asp Gln Gln Val Val Ile Asn
50           55           60
Tyr Ser Ile Val Lys Gly Leu Lys Tyr Asn Gln Ala Thr Pro Thr Phe
65           70           75           80
His Gln Trp Arg Asp Ala Arg Gln Val Tyr Gly Leu Asn Phe Ala Ser
85           90           95
Lys Glu Glu Ala Thr Thr Phe Ser Asn Ala Met Leu Phe Ala Leu Asn
100          105          110
Ile Met Asn Ser Gln Glu Gly Gly Pro Ser Ser Gln Arg Gln Val Gln
115          120          125
Asn Gly Pro Ser Pro Asp Glu Met Asp Ile Gln Arg Arg Gln Val Met
130          135          140
Glu Gln His Gln Gln Gln Arg Gln Glu Ser Leu Glu Arg Arg Thr Ser
145          150          155          160
Ala Thr Gly Pro Ile Leu Pro Pro Gly His Pro Ser Ser Ala Ala Ser
165          170          175
Ala Pro Val Ser Cys Ser Gly Pro Pro Pro Pro Pro Pro Pro Leu Val
180          185          190
Pro Pro Pro Pro Thr Gly Ala Thr Pro Pro Pro Pro Pro Pro Leu Pro

```

-continued

195	200	205
Ala Gly Gly Ala Gln Gly Ser Ser His Asp Glu Ser Ser Met Ser Gly		
210	215	220
Leu Ala Ala Ala Ile Ala Gly Ala Lys Leu Arg Arg Val Gln Arg Pro		
225	230	235 240
Glu Asp Ala Ser Gly Gly Ser Ser Pro Ser Gly Thr Ser Lys Ser Asp		
245	250	255
Ala Asn Arg Ala Ser Ser Gly Gly Gly Gly Gly Leu Met Glu Glu		
260	265	270
Met Asn Lys Leu Leu Ala Lys Arg Arg Lys Ala Ala Ser Gln Ser Asp		
275	280	285
Lys Pro Ala Glu Lys Lys Glu Asp Glu Ser Gln Met Glu Asp Pro Ser		
290	295	300
Thr Ser Pro Ser Pro Gly Thr Arg Ala Ala Ser Gln Pro Pro Asn Ser		
305	310	315 320
Ser Glu Ala Gly Arg Lys Pro Trp Glu Arg Ser Asn Ser Val Glu Lys		
325	330	335
Pro Val Ser Ser Ile Leu Ser Arg Thr Pro Ser Val Ala Lys Ser Pro		
340	345	350
Glu Ala Lys Ser Pro Leu Gln Ser Gln Pro His Ser Arg Met Lys Pro		
355	360	365
Ala Gly Ser Val Asn Asp Met Ala Leu Asp Ala Phe Asp Leu Asp Arg		
370	375	380
Met Lys Gln Glu Ile Leu Glu Glu Val Val Arg Glu Leu His Lys Val		
385	390	395 400
Lys Glu Glu Ile Ile Asp Ala Ile Arg Gln Glu Leu Ser Gly Ile Ser		
405	410	415
Thr Thr		

## (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1889 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: HEAONOT03
- (B) CLONE: 3089412

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TTTAAGTAGG CTATAAAAT CAAGTTGCTG TCTTCAGAGG GTCTGTGGTC CTCTGATCAA	60
CATAGGCTGG TGGGAGTACA GGACTCGCCT CCTCAGGGTT CCCTGTGCTG CCACTTTTCA	120
GCCATGGCCA CAAGTGAACA GAGTATCTGC CAAGCCCGGG CTTCCTGAT GGTCTACGAT	180
GACACCAGTA AGAAATGGGT ACCAATCAAA CCTGGCCAGC AGGGATTGAG CCGGATCAAC	240
ATCTACCACA AACTGCCAG CAACACCTTC AGAGTCGTTG GAGTCAAGTT GCAGGATCAG	300
CAGGTTGTGA TCAATTATTC AATCGTGAAA GGGCTGAAGT ACAATCAGGC CACGCCAACC	360
TTCCACCACT GGCAGATGC CCGCCAGGTC TACGGCTTAA ACTTTGCAAG TAAAGAAGAG	420
GCAACCACGT TCTCAATGC AATGCTGTTT GCCCTGAACA TCATGAATTC CCAAGAAGGA	480
GGCCCTCCA GCCAGCGTCA GGTGCAGAA GGCCCTCTC CTGATGAGAT GGACATCCAG	540
AGAAGACAAG TGATGGAGCA GCACCAGCAG CAGCGTCAGG AATCTCTAGA AAGAAGAACC	600
TCGGCCACAG GGCCATCCT CCCACCAGGA CATCCTTCAT CTGCAGCCAG CGCCCCCGTC	660



-continued

```

TCATGTAGTG GGCCTCCACC GCCCCCCCCA CCTCTAGTCC CACCTCCACC CACTGGGGCT 720
ACCCACCTC CCCCACCCC ACTGCCAGCC GGAGGAGCCC AGGGGTCCAG CCACGACGAG 780
AGCTCCATGT CAGGACTGGC CGCTGCCATA GCTGGGGCCA AGCTGAGAAG AGTCCAACGG 840
CCAGAAGACG CATCTGGAGG CTCCAGTCCC AGTGGGACCT CAAAGTCCGA TGCCAACCGG 900
GCAAGCAGCG GGGGTGGCGG AGGAGGCCTC ATGGAGGAAA TGAACAAACT GCTGGCCAAG 960
AGGAGAAAAG CAGCCTCCCA GTCAGACAAG CCAGCCGAGA AGAAGGAAGA TGAAAGCCAA 1020
ATGGAAGATC CTAGTACCTC CCCCTCTCCG GGGACCCGAG CAGCCAGCCA GCCACCTAAC 1080
TCCTCAGAGG CTGGCCGGAA GCCCTGGGAG CGGAGCAACT CGTGGAGAA GCCTGTGTCC 1140
TCGATTCTGT CCAGAACCCC GTCTGTGGCA AAGAGCCCCG AAGCTAAGAG CCCCCTTCAG 1200
TCGCAGCCTC ACTCTAGGAT GAAGCCTGCT GGGAGCGTGA ATGACATGGC CCTGGATGCC 1260
TTCGACTTGG ACCGGATGAA GCAGGAGATC CTAGAGGAGG TGTGAGAGA GCTCCACAAG 1320
GTGAAGGAGG AGATCATCGA CGCCATCAGG CAGGAGCTGA GTGGGATCAG CACCACGTAA 1380
GGGGCCGGCC TCGCTGCGCT GATTCTGCGA GCCCATCCGG CGACAGAGGA CAGCCAGAAG 1440
CCCAGCCAGC CCCAGACTCC AGTGCACCAG AGCACGCACA GGAGCCTGGG CGCGCTGCTG 1500
TGAAACGTCC TGACCTGTGA TCACACATGA CAGTGAGGAA ACCAAGTGCA ACTCCTGGGT 1560
TTTTTTTAGA TTCTGCTGA CACGGAACAC CAGGTCTGCT CGTCTTTTTT GTGTTTTATA 1620
TTTGCTTATT TAAGGTACAT TTCTTTGGGT TTCTAGAGAC GCCCCTAAGT CACCTGCTTC 1680
ATTAGACGGT TTCCAGGTTT TCTCCAGGT GACGCTGTTA CGCCTCAGC TGGCGGTGAC 1740
AGCCGGCCCA GCGTGGCGCC ACCACACACC GCAGAGCTGT CCAGGCACAG CTCCTCCCC 1800
AGCGCTCATG GTGTTGAAAC TGTCTGTCAT GCACCACGGT GTCTGTGTCC ACACAGTAAT 1860
AAACGGTTTA CTGTCCGCAA AAAAAAAAAA 1889

```

## (2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 393 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: GenBank
- (B) CLONE: 1644453

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```

Met Ser Glu Gln Ser Ile Cys Gln Ala Arg Ala Ser Val Met Val Tyr
 1           5           10           15
Asp Asp Thr Ser Lys Lys Trp Val Pro Ile Lys Pro Gly Gln Gln Gly
 20           25           30
Phe Ser Arg Ile Asn Ile Tyr His Asn Thr Ala Ser Ser Thr Phe Arg
 35           40           45
Val Val Gly Val Lys Leu Gln Asp Gln Gln Val Val Ile Asn Tyr Ser
 50           55           60
Ile Val Lys Gly Leu Lys Tyr Asn Gln Ala Thr Pro Thr Phe His Gln
 65           70           75           80
Trp Arg Asp Ala Arg Gln Val Tyr Gly Leu Asn Phe Ala Ser Lys Glu
 85           90           95
Glu Ala Thr Thr Phe Ser Asn Ala Met Leu Phe Ala Leu Asn Ile Met
100           105           110

```

-continued

```

Asn Ser Gln Glu Gly Gly Pro Ser Thr Gln Arg Gln Val Gln Asn Gly
 115                               120               125

Pro Ser Pro Glu Glu Met Asp Ile Gln Arg Arg Gln Val Met Glu Gln
 130                               135               140

Gln His Arg Gln Glu Ser Leu Glu Arg Arg Ile Ser Ala Thr Gly Pro
 145                               150               155               160

Ile Leu Pro Pro Gly His Pro Ser Ser Ala Ala Ser Thr Thr Leu Ser
 165                               170               175

Cys Ser Gly Pro Pro Pro Pro Pro Pro Pro Pro Val Pro Pro Pro Pro
 180                               185               190

Thr Gly Ser Thr Pro Pro Pro Pro Pro Pro Pro Leu Pro Ala Gly Gly Ala
 195                               200               205

Gln Gly Thr Asn His Asp Glu Ser Ser Ala Ser Gly Leu Ala Ala Ala
 210                               215               220

Leu Ala Gly Ala Lys Leu Arg Arg Val Gln Arg Pro Glu Asp Ala Ser
 225                               230               235               240

Gly Gly Ser Ser Pro Ser Gly Thr Ser Lys Ser Asp Ala Asn Arg Ala
 245                               250               255

Ser Ser Gly Gly Gly Gly Gly Gly Leu Met Glu Glu Met Asn Lys Leu
 260                               265               270

Leu Ala Lys Arg Arg Lys Ala Ala Ser Gln Thr Asp Lys Pro Ala Asp
 275                               280               285

Arg Lys Glu Asp Glu Ser Gln Thr Glu Asp Pro Ser Thr Ser Pro Ser
 290                               295               300

Pro Gly Thr Arg Ala Thr Ser Gln Pro Pro Asn Ser Ser Glu Ala Gly
 305                               310               315               320

Arg Lys Pro Trp Glu Arg Ser Asn Ser Val Glu Lys Pro Val Ser Ser
 325                               330               335

Leu Leu Ser Arg Val Lys Pro Ala Gly Ser Val Asn Asp Val Gly Leu
 340                               345               350

Asp Ala Leu Asp Leu Asp Arg Met Lys Gln Glu Ile Leu Glu Glu Val
 355                               360               365

Val Arg Glu Leu His Lys Val Lys Glu Glu Ile Ile Asp Ala Ile Arg
 370                               375               380

Gln Glu Leu Ser Gly Ile Ser Thr Thr
 385                               390

```

## (2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 380 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:
- (A) LIBRARY: GenBank
  - (B) CLONE: 624964

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

Met Ser Glu Thr Val Ile Cys Ser Ser Arg Ala Thr Val Met Leu Tyr
 1                               5               10               15

Asp Asp Gly Asn Lys Arg Trp Leu Pro Ala Gly Thr Gly Pro Gln Ala
 20                               25               30

Phe Ser Arg Val Gln Ile Tyr His Asn Pro Thr Ala Asn Ser Phe Arg
 35                               40               45

Val Val Gly Arg Lys Met Gln Pro Asp Gln Gln Val Val Ile Asn Cys

```

-continued

50	55	60
Ala Ile Val Arg Gly Val Lys Tyr Asn Gln Ala Thr Pro Asn Phe His		
65	70	75 80
Gln Trp Arg Asp Ala Arg Gln Val Trp Gly Leu Asn Phe Gly Ser Lys		
	85	90 95
Glu Asp Ala Ala Gln Phe Ala Ala Gly Met Ala Ser Ala Leu Glu Ala		
	100	105 110
Leu Glu Gly Gly Gly Pro Pro Pro Pro Ala Leu Pro Thr Trp Ser		
	115	120 125
Val Pro Asn Gly Pro Ser Pro Glu Glu Val Glu Gln Gln Lys Arg Gln		
	130	135 140
Gln Pro Gly Pro Ser Glu His Ile Glu Arg Arg Val Ser Asn Ala Gly		
	145	150 155 160
Gly Pro Pro Ala Pro Pro Ala Gly Gly Pro Pro Pro Pro Gly Pro		
	165	170 175
Pro Pro Pro Pro Gly Pro Pro Pro Pro Gly Leu Pro Pro Ser Gly		
	180	185 190
Val Pro Ala Ala Ala His Gly Ala Gly Gly Gly Pro Pro Pro Ala Pro		
	195	200 205
Pro Leu Pro Ala Ala Gln Gly Pro Gly Gly Gly Gly Ala Gly Ala Pro		
	210	215 220
Gly Leu Ala Ala Ala Ile Ala Gly Ala Lys Leu Arg Lys Val Ser Lys		
	225	230 235 240
Gln Glu Glu Ala Ser Gly Gly Pro Thr Ala Pro Lys Ala Glu Ser Gly		
	245	250 255
Arg Ser Gly Gly Gly Gly Leu Met Glu Glu Met Asn Ala Met Leu Ala		
	260	265 270
Arg Arg Arg Lys Ala Thr Gln Val Gly Glu Lys Thr Pro Lys Asp Glu		
	275	280 285
Ser Ala Asn Gln Glu Glu Pro Glu Ala Arg Val Pro Ala Gln Ser Glu		
	290	295 300
Ser Val Arg Arg Pro Trp Glu Lys Asn Ser Thr Thr Leu Pro Arg Met		
	305	310 315 320
Lys Ser Ser Ser Ser Val Thr Thr Ser Glu Thr Gln Pro Cys Thr Pro		
	325	330 335
Ser Ser Ser Asp Tyr Ser Asp Leu Gln Arg Val Lys Gln Glu Leu Leu		
	340	345 350
Glu Glu Val Lys Lys Glu Leu Gln Lys Val Lys Glu Glu Ile Ile Glu		
	355	360 365
Ala Phe Val Gln Glu Leu Arg Lys Arg Gly Ser Pro		
	370	375 380

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of
  - a) an amino acid sequence of SEQ ID NO:1,
  - b) a naturally-occurring amino acid sequence having at least 95% sequence identity to the sequence of SEQ ID

- 55 NO:1, wherein said amino acid sequence encodes a polypeptide that binds the neural protein FE65.
2. A composition comprising a polypeptide of claim 1 in conjunction with a suitable pharmaceutical carrier.

\* \* \* \* \*